Executive Steering Committee
For A.C.E. Policy II
(ESCAP II)
Report 12
October 11, 2001

ESCAP II: Analysis of Missing Data Alternatives for the Accuracy and Coverage Evaluation

Don Keathley, Anne Kearney,
William Bell
U.S. Census Bureau

U S C E N S U S B U R E A U
Helping You Make Informed Decisions

ANALYSIS OF MISSING DATA ALTERNATIVES FOR THE 2000 A.C.E.

EXECUTIVE SUMMARY

What is the impact of using alternative procedures to account for missing data in the 2000 Accuracy and Coverage Evaluation (A.C.E.)?

At the national level, the standard deviation that we observed from using ignorable missingness[1] alternative missing data procedures (384,115) was approximately the same magnitude as the dual system estimate (DSE) sampling error (378,222). This result increases the level of uncertainty about the DSE.

In 2000, dual system estimates (DSE) were calculated using Census and Accuracy and Coverage Evaluation (A.C.E.) data. Some of the A.C.E. data that were needed to calculate DSEs were missing due to either non-interviews or item non-response. These missing data were accounted for by a set of missing data procedures. These missing data procedures consisted of:

- a non-interview household adjustment (P-Sample only)
- demographic characteristic imputations for race, ethnicity, tenure, sex, and age (P-Sample only)
- probability imputations for correct enumeration (E-Sample) and match, and residency status (P-Sample).

The non-interview household adjustment spreads non-interviewed household weights over interviewed households within non-interview adjustment cells. The characteristic imputation uses national distributions and hot decks to impute for the missing demographic characteristics. The probability imputation uses weighted ratios to impute for unresolved enumeration, match, and resident status.

Alternative procedures could have been used to account for these missing data. Using these alternative procedures would have resulted in different DSEs. We wanted to determine how much variation would result from the use of alternative missing data procedures. In turn, we could then incorporate this estimate of variation into total error and loss function analysis. We also wanted to examine the viability of the non-ignorable missingness procedures for enumeration and resident status (procedures 5 and 7 below). We did this by comparing A.C.E. and Measurement Error Reinterview (MER) results for persons who had an imputed enumeration or resident probability, or both, in the A.C.E. with an observed enumeration or resident status, or both, in the MER.

---

[1] Ignorable missingness assumes that, when conditioned on certain information, non-respondents behave like respondents.

We chose the following seven alternative missing data procedures for our analysis:

1. alternative non-interview adjustment cell definitions
2. a nearest-neighbor non-interview adjustment (adding a non-interviewed household weight to only one interviewed household)
3. the use of late-arriving resolved data only in the missing data procedures
4. the use of logistic regression models for the probability imputations
5. the application of non-ignorable missingness to enumeration status (a lowering of enumeration probabilities)
6. the application of non-ignorable missingness to match status (a lowering of match probabilities)
7. the application of non-ignorable missingness to resident status (a lowering of resident probabilities)

There were 128 possible combinations of the seven alternative missing data procedures above; we calculated DSEs for all 128.

We observed the following from our analysis:

- Non-sampling variability from the use of alternative missing data procedures is considerable. At the national level, the overall magnitude of the standard deviation from using all combinations of the alternative missing data procedures (531,751) is higher than the DSE sampling error (378,222). When we excluded the non-ignorable missingness [2] procedures from the analysis, we observed a standard deviation (384,115) that was approximately the same magnitude as the DSE sampling error.

- There is no evidence to suggest that the non-ignorable missingness procedures that we considered are or are not viable alternative missing data procedures. In our non-ignorable missingness procedures, we lowered correct enumeration, match, and resident probabilities for persons who had a corresponding unresolved status (enumeration, match, and/or resident, respectively).

- The alternative procedures above tended to effect DSEs differently:

    - Procedures 1 and 6 tended to increase the DSEs
    - Procedures 2 and 3 had no apparent effect on the DSEs
    - Procedures 4, 5, 7, and 3 and 4 combined tended to decrease the DSEs

[2] Non-ignorable missingness assumes that, even when conditioned on certain information, non-respondents behave differently than respondents.

1. BACKGROUND

The Accuracy and Coverage Evaluation (A.C.E.) Survey relies on dual system estimation to estimate coverage in Census 2000. The A.C.E. computes dual system estimates (DSE) at the post-stratum level. Post-stratum level DSEs can then be added to form higher level estimates. See Griffin (2000) for details on dual system estimation.

As in most surveys, missing data in the A.C.E. result from non-interviews and item non-response. The A.C.E. must account for these missing data to calculate the DSEs. It does this by implementing a set of missing data procedures to account for the missing data. These data include:

- Noninterviews for P-Sample households
- Interviews with some or all of the following:
    - missing demographic characteristics (race, ethnicity, sex, age, tenure) for P-Sample persons - imputation for E-Sample persons wasn't necessary
    - unresolved match and resident status for P-Sample persons
    - unresolved enumeration status for E-Sample persons

The A.C.E. accounts for these missing data in various ways. It spreads non-interviewed household weights over P-Sample interviewed households. It uses national distributions and hot decks to impute for the missing demographic characteristics. Finally, it uses an imputation cell procedure to impute missing resident, match, and enumeration status probabilities. See Attachment A for a more detailed summary of these missing data procedures.

2. METHODS

We wanted to examine the spread of DSEs when alternative missing data procedures were used on the same A.C.E. data. The resulting range would give us an indication of how sensitive the DSEs are to changes in one or more of the missing data procedures. We settled on the following seven alternatives.

Table 1. Alternative Missing Data Procedures

Alternative Procedure	Description	Motivation for using the Alternative Procedure
Alternative Non-interview Adjustment (NIA) Cell Definitions	Use different NIA cells[1]. These alternative cells were defined on type of basic address, race/ethnicity/tenure, census division, state within division, type of enumeration area, and household size (see Attachment B for a more detailed summary).	Household characteristics may be more homogeneous within alternative NIA cells.
Nearest-Neighbor NIA Procedure	Add each non-interviewed household's (donor) weight to the nearest (in a specified sort) interviewed household's (donee) weight. Each donee would receive no more than one donor's weight.	More homogeneity may result between donor and donee household characteristics when compared to spreading weights over many interviewed households.
Late Data	Assign non-interviewed household weights to late-arriving household interviews only; use the same late-arriving interview information in imputing for probabilities (see Attachment B for a more detailed summary).	Late-arriving interview data may more accurately reflect non-interviews and persons with unresolved match, resident, and enumeration status
Logistic Regression	Assign resident, match, and enumeration probabilities to unresolved cases using a logistic regression model (see Attachment C for a more detailed summary).	Logistic regression models are accepted methods for estimating probabilities. This is what was done in 1990.
Non-ignorable Missingness for Enumeration Probability	Lower the imputed enumeration, match, and resident probabilities for the corresponding unresolved cases. See Attachment D for a discussion on the research and procedure for lowering the probabilities.	Enumeration, match, and resident rates using resolved-only cases may overstate the corresponding rates for unresolved cases.
Non-ignorable Missingness for Match Probability		
Non-ignorable Missingness for Resident Probability		

Note that we didn't include any demographic characteristic imputation alternatives. Our thinking was that the estimates of A.C.E. sampling variance would account for the variation associated with these imputations, with some minor adjustments to the variance procedures, if necessary.

Every alternative procedure contains two levels. One level is using the alternative procedure, the other level is using the procedure used in production. So, we have $2^7 = 128$ alternative procedure combinations (combinations).

[1] Production non-interview adjustment (NIA) cells were defined on block cluster, type of basic address category, recoded A.C.E. sample stratum, and state (Cantwell (2001)).

We grouped these 128 combinations into the following four alternative groups:

Table 2. Alternative Group Descriptions

Alternative Group	Description*	Number of Alternatives
1	AC, NN, LR, LD no non-ignorable missingness	16
2	AC, NN, LR, LD, non-ignorable missingness for all three probabilities	16
3	AC, NN, LR, non-ignorable missingness for either one or two probabilities, no late data combinations	48
4	AC, NN, LR, LD, non-ignorable missingness for either one or two probabilities, late data combinations only	48

* where AC = alternative NIA cell definitions
NN = nearest neighbor imputation
LD = late data
LR = logistic regression

Alternative group 1 (which includes the all-production missing data combination) makes an assumption that the missing data mechanism is ignorable. Ignorability assumes that, conditional on certain information, nonrespondents behave like respondents. For example, the all-production combination assumes nonrespondents behave like respondents within the same imputation cell. In another example, combinations using late data assume nonrespondents behave like the last 30% of respondents. And so on. This type of ignorability assumption is standard in survey estimation.

The true nature of nonresponse may not be ignorable, however: it may be non-ignorable. We took possible non-ignorable missingness into account by lowering the correct enumeration, match, and resident probabilities for all persons with imputed enumeration, match, and resident status, respectively. We did this for all 16 combinations in alternative group 1. The result was an additional 112 alternative procedure combinations. Alternative groups 2-4 above represent these additional combinations. We based the procedure for lowering the imputed probabilities on evaluations of the 1990 Post Enumeration Survey (PES) missing data procedures (see Attachment D).

All of the combinations in alternative group 1 were readily implementable into the A.C.E. We did not share this notion with alternative groups 2-4, however. We did not think the use of the non-ignorable missingness procedures in the A.C.E. were defensible given the basis for their derivation (the 1990 PES) and how we were using this derivation to adjust the 2000 A.C.E. imputed probabilities. Instead, we chose the combinations in alternative groups 2-4 to obtain some measure of variation about the choice of missing

data combinations which include plausible non-ignorability procedures.

Our notion was that the true correct enumeration, match, and resident rates for persons with the respective unresolved status might be lower than the same rates for persons with resolved status; the A.C.E. used data from persons with resolved enumeration, match, and resident status to compute the respective imputed probabilities. Hence, we defined our non-ignorability adjustments as the lowering of imputed probabilities.

Prior to making the non-ignorability adjustments, we knew that

- lowering the imputed correct enumeration probabilities would lower the DSEs,
- lowering the imputed match rates would raise the DSEs,
- lowering the imputed resident rates would lower the DSEs

The bullets above imply that there is a offsetting effect on DSEs, to some degree, for combinations that include non-ignorable missingness for match status and non-ignorable missingness for enumeration status, resident status, or both. The combinations that include the application of all three non-ignorable missingness procedures are in alternative group 2. The remaining combinations are in alternative groups 3 and 4.

Conversely, the bullets also imply that the effect on DSEs could reach a maximum for combinations that include non-ignorable missingness for match status or non-ignorable missingness for enumeration status, resident status, or both. All of these combinations are in alternative groups 3 and 4.

From Table 2, one can see that the difference between alternative groups 3 and 4 is that alternative group 3 does not include any of the late data combinations. We conducted a preliminary application of the alternative missing data procedures. From this preliminary application, we observed that the interaction of the late data procedure with the application of exactly one or two non-ignorable missingness procedures resulted in the largest effects on DSEs.

So, based on the preceding, as we go from alternative group 1 to alternative group 4, we expect successively more variation in the DSEs.

3. RESULTS

We ran the missing data system starting with alternative group (AG) 1 and progressing to AG 4. Attachment E shows tables and charts that depict the resulting DSEs. Tables 1, 2, and 3 of Attachment E provide legends for understanding the subsequent charts and tables. Some things that stand out in Attachment E are:

A. AG 1 vs. AG 2

- The range of DSEs between AGs 1 and 2 were similar (1,266,317.34 and 1,300,959.23, respectively) - Tables 4 and 5
- AG 1 DSEs for each AC-NN-LD-LR combination were larger than the corresponding AG 2 DSEs - Chart 1
- There was little variation in the differences in DSEs between AGs 1 and 2 across the AC-NN-LD-LR combinations

B. AG 3

- The range of DSEs is 1,750,773.05, 484,455.71 and 449,813.72 larger than the ranges for AGs 1 and 2, respectively - Table 6
- Applying non-ignorable missingness to match status only resulted in the largest DSEs for each AC-NN-LR combination - Chart 2
- Applying non-ignorable missingness to resident and enumeration status resulted in the lowest DSEs for each AC-NN-LR combination - Chart 2

C. AG 4

- The range of DSEs is 2,628,487.66, the largest range among the AGs - Table 7
- Applying non-ignorable missingness to match status only resulted in the largest DSEs for each AC-NN-LD-LR combination - Chart 3
- Applying non-ignorable missingness to resident and enumeration status only resulted in the lowest DSEs for each AC-NN-LD-LR combination - Chart 3

D. AGs 1-4

We made the following observations:

- alternative NIA cell definitions
  - increased the DSEs, except when combined with both late data and logistic regression (combination 13 in Graph 4)
  - produced the highest DSEs when combined with both nearest neighbor imputation and late data (combination 11)
- nearest neighbor imputation - no apparent effect on the DSEs
- late data - no apparent effect on the DSEs
- logistic regression - tended to decrease the DSEs
- non-ignorable missingness for enumeration status - decreased the DSEs
- non-ignorable missingness for match status - increased the DSEs
- non-ignorable missingness for resident status - decreased the DSEs
- The tandem of late data and logistic regression:
  - decreased the DSE
  - resulted in the lowest DSEs when taken by themselves (combination 10)

We also investigated whether the non-ignorable missingness procedures we considered are viable. To do this, we examined persons with an unresolved enumeration or resident status, or both, in the A.C.E. and a respective resolved status in the Measurement Error Reinterview (MER). We compared these persons' overall A.C.E. imputed correct enumeration and resident rates to their overall resolved MER rates[2]. The imputed correct enumeration probability for these persons was 0.767; the MER resolved correct enumeration rate for these same persons was 0.754. We thought that this difference (0.013) was small enough to not show any evidence of non-ignorable missingness. The imputed resident rate for these persons (excluding new inmovers) was 0.704; the MER resident rate for these same persons was 0.828. This difference (-0.124) showed possible evidence of non-ignorable missingness in the opposite direction of our study.

However, the outcomes above may have differed had we been able to resolve more persons' enumeration and resident statuses in the MER. This is because the persons with still-unresolved enumeration and resident statuses in the MER may behave differently from the persons whose statuses we did resolve.

4. CONCLUSIONS

- Non-sampling variability from the use of alternative missing data procedures is considerable. At the national level, the overall magnitude of the standard deviation from using all combinations of the alternative missing data procedures (531,751) is higher than the DSE sampling error (378,222). When we excluded the non-ignorable missingness procedures from the analysis, we observed a standard deviation (384,115) that was approximately the same magnitude as the DSE sampling error.

- There is no evidence to suggest that the non-ignorable missingness procedures that we considered are or are not viable alternative missing data procedures.

5. REFERENCES

Belin, Tom R. (2001), *Evaluation of Unresolved Enumeration Status in 2000 Census Accuracy and Coverage Evaluation Program.* Document prepared by Datametrics, Inc. for the Bureau of the Census.

Cantwell, Patrick J. (2001), *Accuracy and Coverage Evaluation Survey: Specifications for the Missing Data Procedures; Revision of Q-25.* Internal Census Bureau Memorandum, DSSD Census 2000 Procedures and Operations Memorandum Q-62.

---

[2] There weren't enough persons with an unresolved match status in the A.C.E. and resolved match status in the MER to do a comparison.

Griffin, Richard (2000), *Accuracy and Coverage Evaluation Survey: Dual System Estimation.* Internal Census Bureau memorandum, DSSD Census 2000 Procedures and Operations Memorandum Series Q-20.

Kearney, Anne T. (2001), *Specifications for the Pseudo-Block Cluster Alternative Cell Definition for the Noninterview Adjustment in Evaluation P1: Evaluation of Bias and Uncertainty Associated with Application of the A.C.E. Missing Data Procedures,* Draft Internal Memorandum.

Spencer, Bruce D. (2000), *Final Report on Methodology for Bias and Uncertainty in Missing Data Procedures.* Document prepared by Abt Associates Inc. and Spencer Statistics, Inc. for the Bureau of the Census.

Summary of Procedures to Address Missing Data in the A.C.E. [3]

I. Introduction

This attachment outlines the missing data procedures used for the Census 2000 Accuracy and Coverage Evaluation (A.C.E.). We applied a noninterview adjustment within cells to account for whole-household nonresponse. A characteristic imputation procedure assigned values for specific missing demographic variables. Depending on the variable, we used hot-deck imputation, imputation from conditional distributions, or a combination of the two. Finally, to people with unresolved resident, match, or enumeration status, we assigned a probability using a method called imputation cell estimation. The probability assigned was based on the status of people in the same imputation cell with resolved status.

For a detailed description of these procedures, see DSSD CENSUS 2000 PROCEDURES AND OPERATIONS MEMORANDUM Q-62, "Accuracy and Coverage Evaluation Survey: Specifications for the Missing Data Procedures; Revision of Q-25."

II. Noninterview Adjustment

Noninterview adjustment was used to address whole-household nonresponse. This procedure was applied only to the P-Sample. Because of our strategy for handling people who moved in or out of sample households, two adjustments were needed: one each based on the household's interview status as of (i) Census Day and (ii) the day of the A.C.E. interview. The two procedures were essentially identical except for the reference date for interview status.

The adjustment cells for noninterview were *block cluster* × *type of basic address*. The three types of basic address were single-family units, apartments, and all others. Generally, weights of noninterviewed units were spread over interviewed units in the same cell. When there were not enough interviews in the cell compared to the number of noninterviews, we spread the weights over a larger category of interviewed units. Note that cells were never collapsed together: weights of noninterviewed units in a problem cell were spread over a broader category, but weights of noninterviewed units in non-problem cells were still spread only within their cell.

III. Characteristic Imputation

There were five variables requiring imputation in the A.C.E.: tenure (owner/non-owner), race, Hispanic origin, age, and sex. We only imputed these characteristics in the P

[3] Pat Cantwell provided this summary.

Sample, as the E-Sample data for these five variables were taken from the Hundred-Percent Census Edited File.

A. Tenure

When tenure was missing, we used the tenure of the nearest previous household with reported tenure and the same type of basic address (single-family unit, apartment, or other).

B. Race and Hispanic origin

For missing race, if there was at least one person in the same household who reported race, we imputed from the distribution in that household. If the whole household was missing race but had at least one person who reported Hispanic origin, then we imputed from the nearest previous household with reported race and the same Hispanic origin. If the whole household was missing both race and Hispanic origin, we imputed from the nearest previous household with reported race. Imputation for missing Hispanic origin was essentially like that for race, with the roles of race and Hispanic origin reversed.

C. Age

The age of persons from single-person households was imputed using the distribution of age in single-person households. In multi-person households, for the spouse, child, parent, or sibling of the reference person, we used a conditional distribution based on the relationship to the reference person and the age of the reference person. For the remaining relationship categories, we used conditional distributions based on the relationship to the reference person.

D. Sex

The sex of persons from single-person households was imputed using the distribution of sex in single-person households. In multi-person households, if sex was reported for only one of the reference person and the spouse (when present in the household), the opposite sex was imputed to the person missing sex. When missing, sex for all remaining persons was imputed using conditional distributions based on the relationship to the reference person.

IV. Unresolved Status

To impute for unresolved match and resident status (for P-Sample people) and enumeration status (for E-Sample people), we used imputation cell estimation. First, all people were placed in cells, according to relevant operational and demographic characteristics. Then each person whose status was unresolved was assigned a

probability based on the status of the resolved people in the same imputation cell.

A. Resident status

For resident status, we used operational characteristics (e.g., the match status of people in the housing unit, whether the person needed follow-up), demographic characteristics (e.g., race, Hispanic origin, tenure), and other variables (e.g., information keyed from follow-up forms) to define imputation cells. Originally, we did not expect to have information from the follow-up activity available for production. A special operation was conducted to key the forms.

B. Match status

To define imputation cells for match status, we used mover status, housing-unit match status (including information on conflicting households), and the presence or absence of imputed data. As most people with unresolved match status had only insufficient information for matching, the variables available for defining imputation cells were severely limited.

C. Correct enumeration status

Here we used operational characteristics (e.g., the match status of people in the housing unit, whether the person needed follow-up), demographic characteristics (e.g., race, Hispanic origin), and other variables (e.g., the presence of imputation in the Census, information keyed from follow-up forms). As with resident status, the follow-up information was not originally expected to be available for use in the missing data procedures.

Alternative NIA Cell and Late Data Procedure Summaries [4]

I. Alternative Cells

The A.C.E. noninterview adjustment cells were defined on (block cluster x type of basic address). There were collapsing rules that used recoded sampling stratum and state (for recoded sampling stratum, the A.C.E. differentiated between small blocks and American Indian blocks on reservations; it also used the demographic-tenure make up of the cluster). We used alternative cells that used Census demographic and geographic data. We matched the P-Sample housing unit file to the Hundred Percent Census Edited File and pulled off census householder information. We took the following variables from the census file (HCEF): race, Hispanic origin (hisp), tenure, type of enumeration area (TEA), and household size. We combined these variables with the following P-Sample variables: type of basic address (toba), division of the United States, and state (there were special codes for nonmatching P-Sample households and P-Sample households that matched to a vacant HCEF household). We recoded and concatenated these variables to form the following pseudo-block clusters: toba || race/hisp/tenure || division || state || TEA || household size. The symbol || represents concatenation. When necessary, we collapsed from right to left (i.e., from household size to TEA, etc.)

II. Late Data Alternative

Some literature proposes that late-arriving completed interviews in the field process are more like the noninterviews than those interviews that we collect in the beginning of the field process. If this were true, an alternative to the production procedure of missing data processing would be to use only later-arriving interviews as donors for the missing data. In the P-Sample, we divided the file into a "late" data file, and the remainder. Late housing units in the P-Sample (used in the whole household noninterview adjustment) were the noninterviews and all late interviews. We defined a housing unit as late if it was in the last 30 percent of the interviews in a Local Census Office. We used the date of the interview to get these. Late persons (used to impute unresolved match and residence status) are all persons with unresolved status and all persons from late housing units.

In the E-Sample, a person was late if he/she came from a housing unit that went to Nonresponse Followup.

---

[4] See Kearney (2001) for details.

Logistic Regression Summary [5]

Logistic regression programs were implemented in SAS. The program for unresolved match status featured 186 predictors including state indicators; characteristics of the sample area including size of the block, type of enumeration area, whether the area was urban or non-urban, and demographic tenure codes capturing predominant ethnic subgroups within the sample areas; individual demographic characteristics including age, ethnicity, gender, tenure (i.e., owner/renter status), and relationship to the reference person in the household; A.C.E. processing characteristics including whether the interview relied on a proxy respondent and, importantly, a "before-follow-up group" variable summarizing the type of match or mismatch between the census and A.C.E. rosters (e.g., match, possible match, partial household non-match, whole-household non-match with housing unit match, whole-household non-match with no housing unit match); and interactions involving these variables. The program for unresolved residence status similarly included 186 predictors. The program for unresolved enumeration status included 202 predictors reflecting a different set of before-follow-up groupings and census processing characteristics such as whether the case was part of the non-response follow-up operation. Each model was fit to cases that went to follow-up.

[5] This attachment is taken from Belin (2001).

Basis for the Procedure for the Non-Ignorable Missingness Procedure [6]

To reflect the impact of varying probabilities on the estimation of a parameter to account for potential nonignorable effects, we simulated a distribution of probabilities that reflected key features of the 1990 Post Enumeration Survey (PES) predicted probabilities. Specifically, we generated values using a beta distribution in such a way as to preserve a mean of 0.773 for 52 cases with probabilities between 0.67 and 1. In particular, we assumed probabilities were dispersed according to 0.67 + 1/3 * betainv(j / 53) for j = 1, 2, ... , 52, where "betainv" produces quantiles for a beta distribution; in this case, we chose a beta distribution with parameters a = 0.312 and b = 0.688 based on the relative difference between 0.773 and the minimum value of 0.67 to the width of the entire interval from 1.0 to 0.67. This procedure reproduced the overall mean of 0.773 and implied that the probabilities ran from 0.670 to 0.999. These probabilities were transformed to a logit scale, and a constant amount was subtracted from each. By trial and error, it was noted that subtracting 0.83 on the logit scale from each of the logit-transformed probabilities and then transforming back to the probability scale would produce an average probability of just below 0.615.

One more step was needed to apply the results of the previous paragraph to the 2000 A.C.E. data. The final step in anchoring a nonignorable model for 2000 A.C.E. data in an empirical framework was based on characterizing how a logistic regression parameter of 0.83 would compare to the magnitude of parameters controlling for observed characteristics in a large logistic regression model. A logistic regression model similar to the 2000 A.C.E. model, with 182 predictors, was fit to 1990 PES data. It was found that 164 parameters were smaller in absolute value than 0.83 and 18 parameters were larger in absolute value than 0.83, putting 0.83 at the 90th percentile of the distribution of the absolute values of logistic regression parameters. In that spirit, nonignorable alternatives on 2000 A.C.E. data were based on identifying a scalar value below which 90% of the logistic regression parameters fall in absolute value, and then alternatively adding that value to each logistic-transformed predicted probability (yielding a higher predicted probability) or subtracting that value from each logistic-transformed predicted probability (yielding a lower predicted probability).

[6] This attachment is taken from Belin (2001).

Table 1. Alternative Group Descriptions

Alternative Group	Description
1	AC, NN, LR, LD no non-ignorable missingness
2	AC, NN, LR, LD, non-ignorable missingness for all three probabilities
3	AC, NN, LR, non-ignorable missingness for either one or two probabilities, no late data combinations
4	AC, NN, LR, LD, non-ignorable missingness for either one or two probabilities, late data combinations only

where AC = alternative NIA cell definitions
NN = nearest neighbor imputation
LD = late data
LR = logistic regression

Table 2. Combination Explanations

Combination Number	Alternatives	Combination Number	Alternatives
1	AC	9	NN, LR
2	NN	10	LD, LR
3	LD	11	AC, NN, LD
4	LR	12	AC, NN, LR
5	AC, NN	13	AC, LD, LR
6	AC, LD	14	NN, LD, LR
7	AC, LR	15	All four
8	NN, LD	16	None

Table 3. Non-Ignorable Missingness Code Explanations

Code	Explanation
blank (None)	No non-ignorable missingness
a (E)	Enumeration status only
b (M)	Match status only
c (R)	Resident status only
d (E,M)	Both enumeration and match status
e (E,R)	Both enumeration and resident status
f (M,R)	Both match and resident status
g (All)	All three

Table 4. Alternative Group 1 DSEs - No Non-Ignorable Missingness (sorted by DSE)

(AC) Alternative NIA Cell Definitions	(NN) Nearest Neighbor NIA	(LD) Late Data	(LR) Logistic Regression	Combination Number	*DSE Range: 1,266,320.34* DSE
		x	x	10	276,451,522.59
	x	x	x	14	276,475,844.86
			x	4	276,801,391.60
x		x	x	13	276,801,620.93
				16 (Production)	**276,848,872.57**
	x		x	9	276,854,850.49
	x			2	276,905,555.55
x	x	x	x	15	276,951,831.77
		x		3	277,282,033.13
	x	x		8	277,308,762.19
x			x	7	277,375,277.57
x				1	277,400,435.12
x	x		x	12	277,502,671.17
x	x			5	277,525,400.92
x		x		6	277,573,136.90
x	x	x		11	277,717,839.93

NOTE: An "x" indicates that the combination used the alternative; a blank indicates that the combination used the production method/definition.

Table 5. Alternative Group 2 DSEs - Non-Ignorable Missingness for *ALL* Probability Imputations (Sorted by DSE)

(AC) Alternative NIA Cell Definitions	(NN) Nearest Neighbor NIA	(LD) Late Data	(LR) Logistic Regression	Combination Number	*DSE Range: 1,300,959.23* DSE
		x	x	10	276,036,646.73
	x	x	x	14	276,062,761.14
x		x	x	13	276,319,558.89
x	x	x	x	15	276,463,946.99
			x	4	276,532,346.72
				16	276,569,860.95
	x		x	9	276,582,225.38
	x			2	276,623,718.75
		x		3	276,971,915.97
	x	x		8	277,000,216.04
x			x	7	277,048,937.71
x				1	277,065,874.26
x	x		x	12	277,173,188.58
x	x			5	277,189,628.56
x		x		6	277,197,690.13
x	x	x		11	277,337,605.96

NOTE: An "x" indicates that the combination used the alternative; a blank indicates that the combination used the production method/definition.

Chart 1: DSEs for Alternative Groups 1 & 2

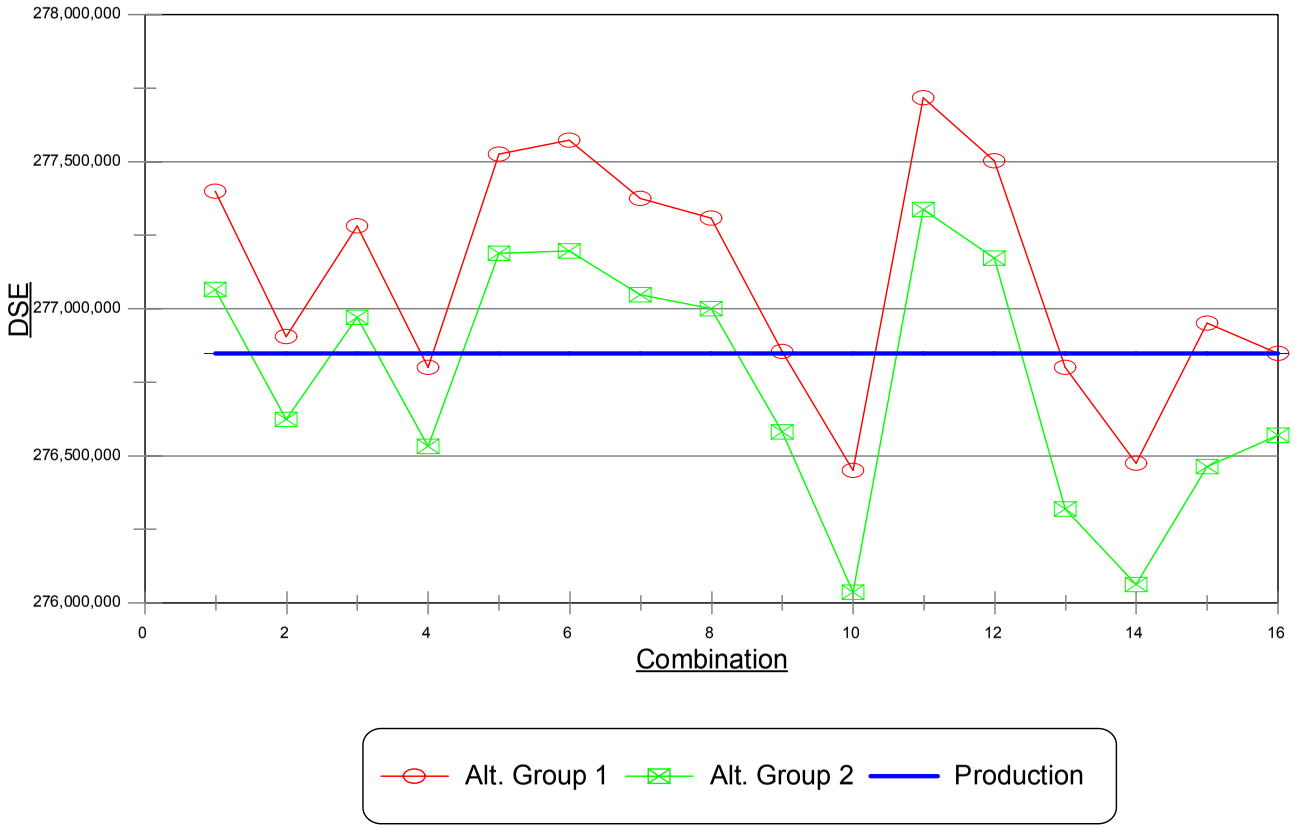

Table 6. Alternative Group 3 DSEs - no late data combinations (sorted by DSE)

(AC) Alternative NIA Cell Definitions	(NN) Nearest Neighbor NIA	(LD) Late Data	(LR) Logistic Regression	(E) Correct Enumer.	(M) Match	(R) Resident	Combination Number	*DSE Range: 1,750,773.05* DSE
				Non-ignorable Missingness				
			x	x		x	4e	276,176,994.90
				x		x	16e	276,191,269.06
	x		x	x		x	9e	276,226,315.15
	x			x		x	2e	276,244,007.52
			x	x			4a	276,416,193.39
				x			16a	276,450,369.84
	x		x	x			9a	276,469,516.91
	x			x			2a	276,506,922.71
			x			x	4c	276,561,660.61
						x	16c	276,589,165.94
	x		x			x	9c	276,611,109.62
	x					x	2c	276,642,027.07
x				x		x	1e	276,740,958.63
x			x	x		x	7e	276,744,487.84
			x	x	x		4d	276,853,079.57
x	x			x		x	5e	276,864,794.99
x	x		x	x		x	12e	276,868,883.14
	x		x	x	x		9d	276,907,007.60
			x		x	x	4f	276,917,698.57
				x	x		16d	276,919,101.42
	x		x		x	x	9f	276,967,708.00
					x	x	16f	276,968,522.79
	x			x	x		2d	276,977,005.02
x			x	x			7a	276,988,905.53
x				x			1a	277,000,764.53
	x				x	x	2f	277,022,505.78
x	x		x	x			12a	277,115,922.87
x	x			x			5a	277,125,337.79
x			x			x	7c	277,130,309.44
x						x	1c	277,140,013.49
			x		x		4b	277,239,108.32
x	x		x			x	12c	277,255,075.68
x	x					x	5c	277,264,241.76
	x		x		x		9b	277,293,174.06
					x		16b	277,318,538.16
x			x	x	x		7d	277,361,918.43
	x				x		2b	277,376,574.50
x				x	x		1d	277,402,318.12
x			x		x	x	7f	277,435,344.94
x					x	x	1f	277,465,585.55
x	x		x	x	x		12d	277,488,807.75
x	x			x	x		5d	277,526,899.17
x	x		x		x	x	12f	277,559,970.54
x	x				x	x	5f	277,589,735.69
x			x		x		7b	277,748,997.20
x					x		1b	277,802,789.14
x	x		x		x		12b	277,876,267.38
x	x				x		5b	277,927,767.95

NOTE: An "x" indicates that the combination used the alternative; a blank indicates that the combination used the production method/definition.

Chart 2: Alternative Group 3 DSEs

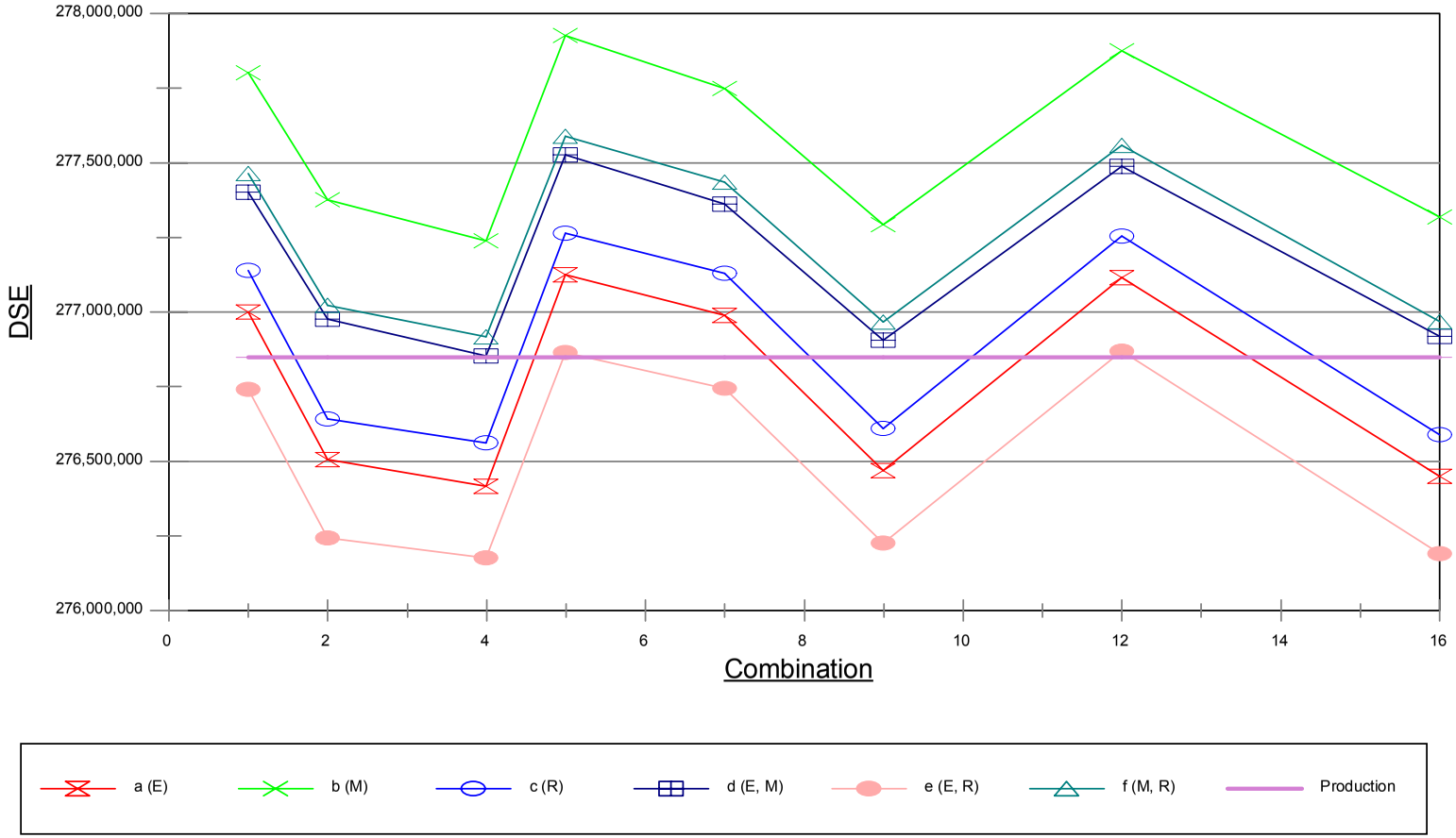

Table 7. Alternative Group 4 DSEs - late data combinations only (sorted by DSE)

(AC) Alternative NIA Cell Definitions	(NN) Nearest Neighbor NIA	(LD) Late Data	(LR) Logistic Regression	(E) Non-ignorable Missingness Correct Enumer.	(M) Non-ignorable Missingness Match	(R) Non-ignorable Missingness Resident	Combination Number	*DSE Range: 2,628,487.66* DSE
		x	x	x		x	10e	275,623,183.75
	x	x	x	x		x	14e	275,648,188.61
		x	x	x			10a	275,863,865.89
	x	x	x	x			14a	275,888,137.93
x		x	x	x		x	13e	275,966,608.17
x	x	x	x	x		x	15e	276,117,060.00
		x	x			x	10c	276,210,062.05
x		x	x	x			13a	276,212,901.94
	x	x	x			x	14c	276,235,123.86
x	x	x	x	x			15a	276,362,541.18
		x	x	x	x		10d	276,379,641.63
	x	x	x	x	x		14d	276,405,260.32
		x		x		x	3e	276,488,450.86
	x	x		x		x	8e	276,515,117.71
x		x	x			x	13c	276,554,527.15
		x	x		x	x	10f	276,624,691.70
	x	x	x		x	x	14f	276,650,864.72
x		x	x	x	x		13d	276,651,399.85
x	x	x	x			x	15c	276,705,553.47
x		x		x		x	6e	276,789,796.64
x	x	x	x	x	x		15d	276,793,908.40
		x		x			3a	276,809,977.57
	x	x		x			8a	276,836,674.48
x		x	x		x	x	13f	276,908,454.04
x	x	x		x		x	11e	276,936,565.86
		x				x	3c	276,959,696.99
		x	x		x		10b	276,968,734.44
	x	x				x	8c	276,986,400.17
	x	x	x		x		14b	276,994,405.14
x	x	x	x		x	x	15f	277,053,408.21
x		x		x			6a	277,100,254.94
x		x	x		x		13b	277,241,316.44
x	x	x		x			11a	277,244,515.02
x		x				x	6c	277,261,888.10
x	x	x	x		x		15b	277,384,386.89
x	x	x				x	11c	277,409,104.74
		x			x	x	3f	277,444,276.07
		x		x	x		3d	277,453,055.44
	x	x			x	x	8f	277,472,615.50
	x	x		x	x		8d	277,481,856.23
x		x		x	x		6d	277,641,304.99
x		x			x	x	6f	277,670,714.56
x	x	x		x	x		11d	277,777,134.01
x	x	x			x	x	11f	277,811,068.59
		x			x		3b	277,926,573.68
	x	x			x		8b	277,955,410.16
x		x			x		6b	278,115,409.89
x	x	x			x		11b	278,251,671.41

NOTE: An "x" indicates that the combination used the alternative; a blank indicates that the combination used the production method/definition.

Chart 3: Alternative Group 4 DSEs

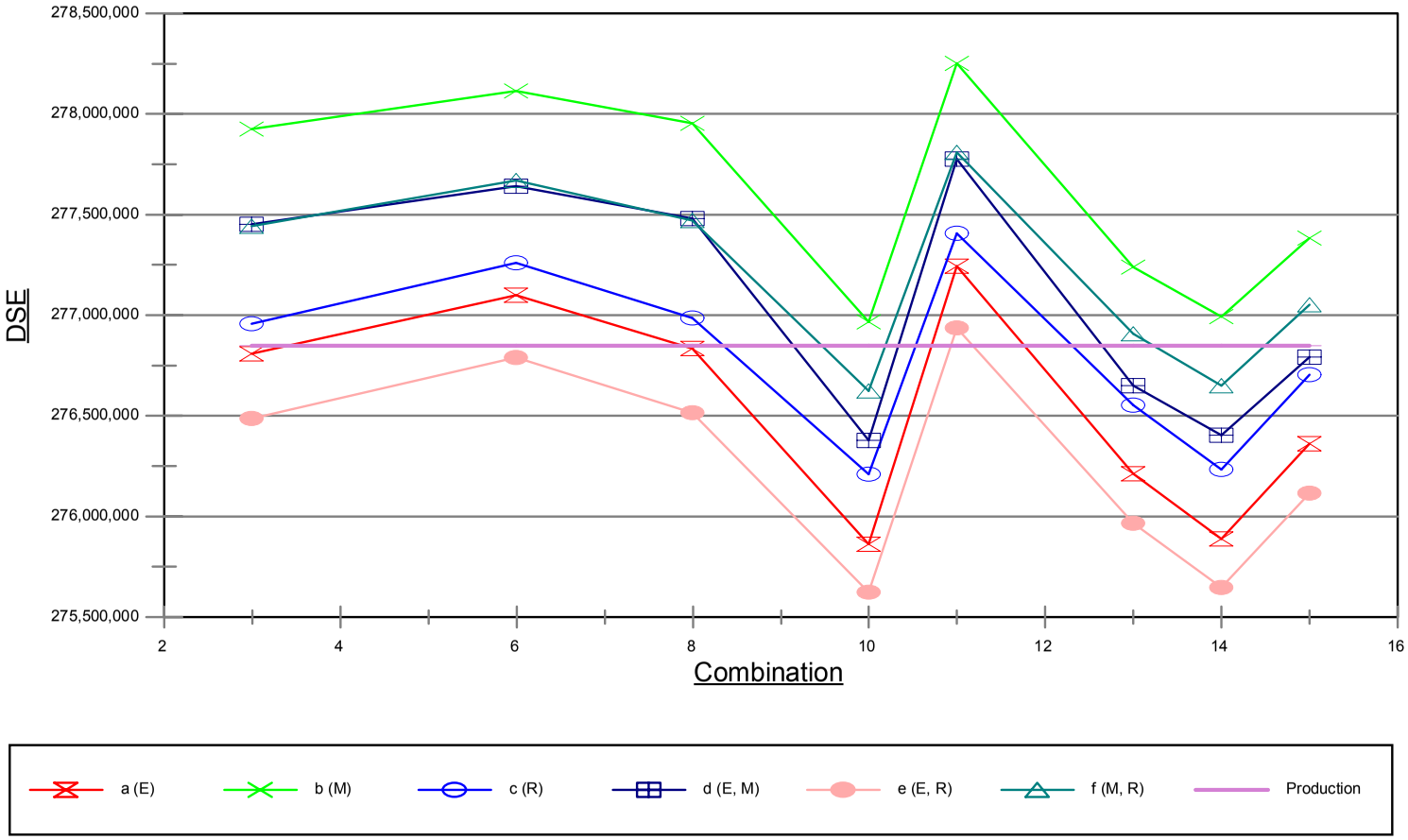

Table 8. All Alternative Group DSEs (sorted by DSE)

(AC) Alternative NIA Cell Definitions	(NN) Nearest Neighbor NIA	(LD) Late Data	(LR) Logistic Regression	(E) Correct Enumer.	(M) Match	(R) Resident	Combination Number	*DSE Range: 2,628,487.66* DSE
				Non-ignorable Missingness				
		x	x	x		x	10e	275,623,183.75
	x	x	x	x		x	14e	275,648,188.61
		x	x	x			10a	275,863,865.89
	x	x	x	x			14a	275,888,137.93
x		x	x	x		x	13e	275,966,608.17
		x	x	x	x	x	10g	276,036,646.73
	x	x	x	x	x	x	14g	276,062,761.14
x	x	x	x	x		x	15e	276,117,060.00
			x	x		x	4e	276,176,994.90
				x		x	16e	276,191,269.06
		x	x			x	10c	276,210,062.05
x		x	x	x			13a	276,212,901.94
	x		x	x		x	9e	276,226,315.15
	x	x	x			x	14c	276,235,123.86
	x			x		x	2e	276,244,007.52
x		x	x	x	x	x	13g	276,319,558.89
x	x	x	x	x			15a	276,362,541.18
		x	x	x	x		10d	276,379,641.63
	x	x	x	x	x		14d	276,405,260.32
			x	x			4a	276,416,193.39
				x			16a	276,450,369.84
		x	x				10	276,451,522.59
x	x	x	x	x	x	x	15g	276,463,946.99
	x		x	x			9a	276,469,516.91
	x	x	x				14	276,475,844.86
		x		x		x	3e	276,488,450.86
	x			x			2a	276,506,922.71
	x	x		x		x	8e	276,515,117.71
			x	x	x	x	4g	276,532,346.72
x		x	x			x	13c	276,554,527.15
			x			x	4c	276,561,660.61
				x	x	x	16g	276,569,860.95
	x		x	x	x	x	9g	276,582,225.38
						x	16c	276,589,165.94
	x		x			x	9c	276,611,109.62
	x			x	x	x	2g	276,623,718.75
		x	x		x	x	10f	276,624,691.70
	x					x	2c	276,642,027.07
	x	x	x		x	x	14f	276,650,864.72
x		x	x	x	x		13d	276,651,399.85
x	x	x	x			x	15c	276,705,553.47
x				x		x	1e	276,740,958.63
x			x	x		x	7e	276,744,487.84
x		x		x		x	6e	276,789,796.64
x	x	x	x	x	x		15d	276,793,908.40
			x				4	276,801,391.60
x		x	x				13	276,801,620.93
		x		x			3a	276,809,977.57

Table 8. All Alternative Group DSEs (sorted by DSE)

(AC) Alternative NIA Cell Definitions	(NN) Nearest Neighbor NIA	(LD) Late Data	(LR) Logistic Regression	(E) Correct Enumer.	(M) Match	(R) Resident	Combination Number	*DSE Range: 2,628,487.66* DSE
				Non-ignorable Missingness				
	x	x		x			8a	276,836,674.48
							16	**276,848,872.57**
			x	x	x		4d	276,853,079.57
	x		x				9	276,854,850.49
x	x			x		x	5e	276,864,794.99
x		x	x	x		x	12e	276,868,883.14
	x						2	276,905,555.55
	x		x	x	x		9d	276,907,007.60
x		x	x		x	x	13f	276,908,454.04
			x		x	x	4f	276,917,698.57
				x	x		16d	276,919,101.42
x	x	x		x		x	11e	276,936,565.86
x	x	x	x				15	276,951,831.77
		x				x	3c	276,959,696.99
	x		x		x	x	9f	276,967,708.00
					x	x	16f	276,968,522.79
		x	x		x		10b	276,968,734.44
		x		x	x	x	3g	276,971,915.97
	x			x	x		2d	276,977,005.02
	x	x				x	8c	276,986,400.17
x			x	x			7a	276,988,905.53
	x	x	x		x		14b	276,994,405.14
	x	x		x	x	x	8g	277,000,216.04
x				x			1a	277,000,764.53
	x				x	x	2f	277,022,505.78
x			x	x	x	x	7g	277,048,937.71
x	x	x	x		x	x	15f	277,053,408.21
x				x	x	x	1g	277,065,874.26
x		x		x			6a	277,100,254.94
x		x	x	x			12a	277,115,922.87
x	x			x			5a	277,125,337.79
x			x			x	7c	277,130,309.44
x						x	1c	277,140,013.49
x	x		x	x	x	x	12g	277,173,188.58
x	x			x	x	x	5g	277,189,628.56
x		x		x	x	x	6g	277,197,690.13
			x		x		4b	277,239,108.32
x		x	x		x		13b	277,241,316.44
x	x	x		x			11a	277,244,515.02
x		x	x			x	12c	277,255,075.68
x		x				x	6c	277,261,888.10
x	x					x	5c	277,264,241.76
		x					3	277,282,033.13
	x		x		x		9b	277,293,174.06
	x	x					8	277,308,762.19
					x		16b	277,318,538.16
x	x	x		x	x	x	11g	277,337,605.96
x			x	x	x		7d	277,361,918.43
x			x				7	277,375,277.57

Table 8. All Alternative Group DSEs (sorted by DSE)

(AC) Alternative NIA Cell Definitions	(NN) Nearest Neighbor NIA	(LD) Late Data	(LR) Logistic Regression	(E) Correct Enumer.	(M) Match	(R) Resident	Combination Number	*DSE Range: 2,628,487.66* DSE
					Non-ignorable Missingness			
	x				x		2b	277,376,574.50
x	x	x	x		x		15b	277,384,386.89
x							1	277,400,435.12
x				x	x		1d	277,402,318.12
x	x	x				x	11c	277,409,104.74
x			x		x	x	7f	277,435,344.94
		x			x	x	3f	277,444,276.07
		x		x	x		3d	277,453,055.44
x					x	x	1f	277,465,585.55
	x	x			x	x	8f	277,472,615.50
	x	x		x	x		8d	277,481,856.23
x		x	x	x	x		12d	277,488,807.75
x	x		x				12	277,502,671.17
x	x						5	277,525,400.92
x	x			x	x		5d	277,526,899.17
x		x	x		x	x	12f	277,559,970.54
x		x					6	277,573,136.90
x	x				x	x	5f	277,589,735.69
x		x		x	x		6d	277,641,304.99
x		x			x	x	6f	277,670,714.56
x	x	x					11	277,717,839.93
x			x		x		7b	277,748,997.20
x	x	x		x	x		11d	277,777,134.01
x					x		1b	277,802,789.14
x	x	x			x	x	11f	277,811,068.59
x		x	x		x		12b	277,876,267.38
		x			x		3b	277,926,573.68
x	x				x		5b	277,927,767.95
	x	x			x		8b	277,955,410.16
x		x			x		6b	278,115,409.89
x	x	x			x		11b	278,251,671.41

NOTE: An "x" indicates that the combination used the alternative; a blank indicates that the combination used the production method/definition.

Chart 4: DSEs for all Combinations